



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

THE USE OF AN OBJECTIVE SCALE FOR GRADING HANDWRITING

HERSCHEL T. MANUEL
Greencastle, Indiana

For the last few years a great deal of attention has been given to the problem of educational measurement. Attention has been called repeatedly to the need for objective bases for the evaluation of school products. As a result of this movement a number of objective scales designed to furnish such bases have been published. These published scales have in turn furnished an opportunity for experimentation and test from which a better understanding of the general problem is emerging. The writer has been engaged in a study of the Ayres Measuring Scale for Handwriting (Russell Sage Foundation, New York City, 1912) from the viewpoint of the variability in the grades assigned to the same writing when the scale is used. It is the purpose of this article to report very briefly some of the main results of this study and their apparent implications for the problem above mentioned.¹

The Ayres scale, it will be remembered, consists of reproductions of samples of school handwriting arranged in order of quality on the basis of legibility and assigned certain numerical values. In all, twenty-four different specimens of handwriting are used. These are arranged in eight groups of three specimens each and printed on a rectangular sheet of heavy paper. Above the groups appear the numbers 20, 30, 40, 90, respectively. These numbers indicate the comparative value of the writing displayed below them. The legibility of the samples was determined by actual timed readings of a number of paid investigators reading a

¹ The writer acknowledges with thanks his indebtedness to a large number of different persons, whom space forbids to mention by name, in the study underlying this article. Special thanks are due to Dr. Frank N. Freeman. For a fuller discussion of the investigation see the Master's essay by the writer in the library of the University of Chicago.

large number of samples of elementary-school writing. The grading of handwriting by the use of the scale is simply a matter of comparison of the sample being graded with the samples on the scale, and the giving of an appropriate numerical value in terms of the values appearing above the specimens on the scale. The directions on the scale are very simple: "To measure the quality of a sample of handwriting slide it along the scale until a writing of corresponding value is found. The number in black at the top of the scale above this represents the value of the writing being measured." For purposes of this discussion the expression "to measure the quality of" may be considered synonymous with the phrase "to grade."

The investigation has consisted of the collection of a fairly large number of handwriting grades assigned by different persons working under various conditions and of a study of the variability among them. It has been the intention to use the grades of such persons as might be expected to use the scale in practical work and yet who have had no training in its use other than independent practice. Neither the number of papers graded nor the number of graders contributing to these results in the different parts of the investigation can be detailed here. It will be sufficient to state that the grades of a total of more than a thousand papers are used and that more than two hundred persons contributed to the grading, grading sets varying from eight papers to one hundred and seventy-six, some with the scale and some without and some by both methods.

Briefly stated, the direct conclusion from the investigation is that miscellaneous graders vary rather widely in their use of the Ayres scale when their only training in its use is independent practice. Table I shows how extreme this variability may be among persons presumably expert in their judgment of handwriting. The table shows the grades of eight samples of writing (numbered 1-8) given by three supervisors of writing using the Ayres scale. The figures at the bottom represent scale values. The figures above are the identification numbers of the samples graded and are placed above the values given. Each row represents the grading of one of the three supervisors. The table

should be read as follows: One judge rated paper No. 2 at 20 on the Ayres scale, a second judge at 40, and the third at 75. The same judges rated paper No. 4 at 20, 30, and 50, respectively.

TABLE I

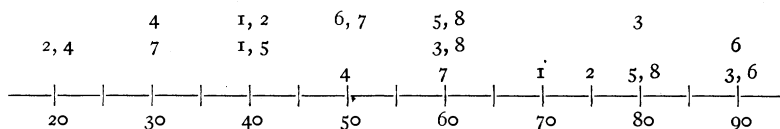


Fig. 1 shows graphically the distribution of values given to paper No. 6 of this series by nineteen members of a class in experimental education after considerable practice with the scale. The heights of the columns represent the percentage of the number in the group (19) assigning the Ayres-scale values indicated below. The second part of the figure gives for comparison the average values of handwriting in the various school grades as proposed by Dr. Frank N. Freeman for tentative standards (*Elementary School Teacher*, XIV, 170).

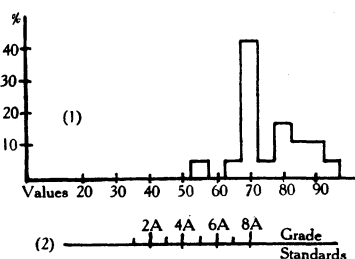


FIG. 1

The characteristic distribution of the grades assigned to separate papers graded in a series is illustrated in Fig. 2. The same eight papers as appeared in Table I are the basis of the grades here represented. The first part of the figure shows the characteristic distribution of values by one hundred eighteen persons, most of whom were relatively unfamiliar with the scale; the second part by a group of nineteen after practice with the scale and after a suggestion of the possibility of using 5's as well as 10's in the grading had been given; the third part by a group of sixty-five grading the papers by the ordinary school method of grading, i.e., on the basis of the advancement of the pupils producing the writing and the passing-grade of the school. The heights of the columns represent, as in Fig. 1, the percentage of the number in the group assigning the values indicated below. The values are indicated in terms of deviations from a middle value. In the first two

parts of the figure the middle value taken is a multiple of ten lying near the median of the distribution of the grades given by the various graders in the group, and in the third part the middle values are multiples of five lying near the median. Suppose, for example, in a group of six graders three give the value 70, two 80, and one 60. The middle value in this distribution would be taken as 70. The table would show then that 50 per cent of the graders assigned a value deviating from the middle value by 0, $33\frac{1}{3}$ per

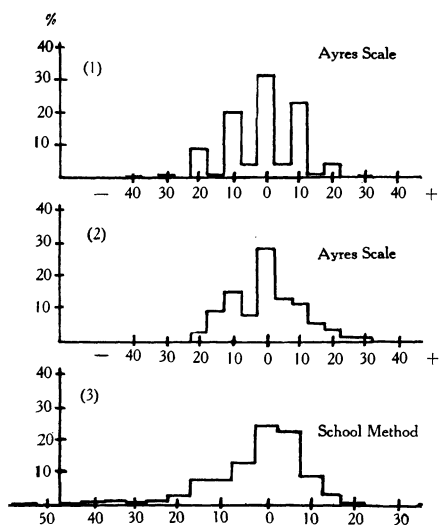


FIG. 2

cent by 10, and $16\frac{2}{3}$ per cent by -10 . It will be observed that the same absolute differences in grades are represented by longer differences on the base line in the third part of the figure than in the other two parts. This change of scale is made to enable one to make a direct comparison of the real significance of the distributions and is made necessary by the fact that the same absolute differences in the two methods of grading form different proportionate parts of the ranges of values used in each.

The curves in Fig. 3 represent the Ayres-scale grades of the writing of two schools of the same city as assigned by two different persons, the same two persons grading each school. The figures at the left indicate the values given to the papers of the school grades indicated below. The grades are based upon from twenty-three to twenty-six papers from each grade. The dotted line represents the grades of one grader and the unbroken line those of the other.

So much by way of illustration; a summary of some of the leading facts brought out by the study follows.

With reference to the variability of the grades of the separate papers of a series it has appeared that:

On the average the chances are more than one in two that two graders after practice in the use of the scale will assign values to the same paper differing by 10 or more, and more than one in six that the values will differ by 20 or more.

The average deviation of the grades of individuals from the group average in grading a single paper with the scale may be as low as 3 or as high as 11.

In general the grades of the same individual at different times may be expected to vary about one-half as much as do the grades of different individuals.

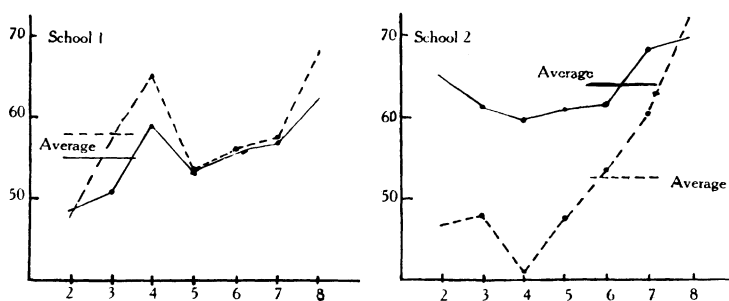


FIG. 3

In general there is a greater variability in the absolute values assigned to writing by the ordinary school method of grading than in those assigned on the basis of the scale. This variability is more marked on the papers that are judged to be below passing.

From the viewpoint of variability the scale appears to have little or no superiority over the other method of grading as a means for finding the relative rank of papers.

With reference to the average grades of series of papers such as are used in comparisons of the writing of groups it has appeared that:

A difference of as much as 10 (a full step on the Ayres scale and twice the difference between the averages of the writing of two successive school grades as shown in Fig. 1) in the averages assigned to a series of twenty-five papers by different persons after practice is common.

In terms of deviations from group averages the variability in the use of the school method is about three halves as great as that

in the use of the scale, and the variability in the use of a subjective scale in which only the zero and one-hundred points are defined almost twice as large as the Ayres-scale variability.

Independent practice in the use of the scale without instruction or exchange of experience may not be expected to reduce the variability if 5's as well as 10's are used in the grading from the first.

The deviation of an individual from a group may be radically modified by experience, but two gradings of the same papers by the same person close to each other in point of time and unmodified by special experience with reference to the grading of handwriting tend to be, from the viewpoint of averages, approximately the same.

The deviation of the average given by a single individual from the group average may be quite different for different sets of papers.

The obtaining of and use of a "personal equation" for correcting the grades of various individual graders that the results of their work may be directly compared require elaborate tests and carefully controlled conditions as to the character and number of the samples graded and the time and experience in grading that intervene between the various gradings used. The evidence at hand does not indicate that such "personal equations" will be generally useful in work with the scale.

In comparisons of the averages of the grades of sets of twenty-five papers graded by different persons any real difference in quality should not be assumed unless the difference in the averages is as much as $7\frac{1}{2}$, three-fourths of a step on the Ayres scale.

In comparisons of the writing of various schools based on the averages of sets containing not to exceed twenty-five papers from each grade from grades two to seven, if the grading is done by different persons using the Ayres scale, any real difference in quality should not be assumed unless the difference in the averages is as much as 5, one-half step on the Ayres scale.

In averages of twenty-five papers given by different groups of four graders each differences less than 5 should not be considered as representing real differences in quality without information as to the characteristic grading of members of the groups.

Such are the facts; what do they mean? The opponent of attempts at exact measurement in education will be inclined to

say, "I told you so! Scales may be all right for the *theorist*, but they have no value for the *practical* man." Another who would be unwilling to go quite so far may say, "No, the fundamental idea of scales is sound, but I never did believe that this particular scale is a good one." Still another who views with grave doubts the ability of the ordinary school man to do anything well may say, "It is merely an evidence of the gross inability of those who did the grading." It would, however, be a long jump indeed to the conclusions just quoted from the real net result of the investigation, viz., that miscellaneous graders vary rather widely in the use of the Ayres scale when their only training in its use is independent practice. The real significance of the results, it appears, is twofold: first, they furnish experimental data for the interpretation of the results when the scale is used under the conditions described in this experiment, and in the second place they furnish a basis for a clearer understanding of the general problem of the scale in educational measurement. The first of these is sufficiently clear from the results themselves; the second will justify some further discussion.

It will hardly be necessary to say anything in support of evaluation as a factor in the work of the schools. Very few, I suppose, would deny that some estimate of the results of educational practice is desirable. Indeed, the absence of all evaluation in school work would set that field off in opposition to all of the rest of human life, which is characterized everywhere by implicit or explicit judgments of value.

Grading is a form of evaluation. In judging the output of any productive system two considerations arise, expressed by the terms quantity and quality. Grading has to do with the determination of and expression of the latter. Quantity is *one* attribute or aspect of a thing. The quality of a thing is the same as the being of the thing itself—it is a complex of *all* the attributes. The accurate expression of quality in quantitative units necessarily involves the quantitative description of each and all of the separate unit factors which constitute the object the quality of which is being determined. The measurement of quality is a measurement of the unit factors which make the object just what it is. The

determination of the quality, therefore, will vary in difficulty with the number of separate factors to be considered and the difficulty of evaluating them.

Measurement consists in a comparison of the thing to be measured with a standard unit of the same thing. To measure, for example, a linear distance we make a comparison of the distance with a unit of distance, and to measure an angle we use a unit angle. Any such distance or angle may, of course, be employed as a unit, but the advantage of having certain well-established standard units is universally recognized in this field, and such units are employed.

Consider now the measurement of the quality of handwriting. It is clear from the analogy that the measurement must be in terms of a unit of quality, and, of course, any quality may be used as a unit for measurement. To carry the analogy a bit farther, just as the meter is described as the distance between certain transverse lines on a certain bar, so a unit of handwriting may be merely some specimen published so that it will be generally available and named so that it may be referred to conveniently. Suppose in the absence of another name we call the unit for the measurement of the quality of handwriting a "scrib." Then in the expression of the quality of a sample of handwriting we are measuring we should say that the quality is, for example, two scribs, meaning that the quality of the sample measured is twice that of the unit sample.

The difficulty with this form of measurement for the quality of handwriting is that the quality of handwriting is not *immediately* quantitative. The quality of handwriting involves a number of factors, including the slant, the alignment, the legibility, the beauty, etc. The use of a unit of handwriting, therefore, necessitates the isolation and comparison of all the separate factors (except as these may be arbitrarily limited by definition). This complexity makes impractical any attempts to evaluate handwriting in terms of the quality of a single specimen such as described. Instead, scales consisting of a number of samples have been devised. The real unit of measurement is the difference in quality between the successive specimens on the scale, but the quality is expressed in terms of the samples themselves by means of certain identification

symbols (such as the 20, 30, etc., of the Ayres scale), rather than directly in terms of the unit.

Now it is evident that an enumeration of all the separate factors of handwriting would be rather difficult. Moreover, as the number of these to be considered in the evaluation of the writing increase, the difficulty of accurate use of a scale increases. There is still a further difficulty in the impossibility of comparison of the various factors by actual physical superposition or even juxtaposition. It is not strange, therefore, that persons without special training vary in the use of the scale. Consider, for example, how complex the use of a red maple rule with a brass edge would be if all of these factors had to be considered in measuring a table! A truer analogy than measurement of dimensions would be that of the grading of agricultural products.

Several conclusions may now be stated:

1. It is clear that generally accepted standards in terms of which the quality of writing may be determined and expressed are desirable.

2. For practical purposes the publication of samples of handwriting is a serviceable part of the description of such standards.

3. The publication of samples must be accompanied by accurate quantitative descriptions as far as possible in terms of the factors to be considered in the grading. A description of the basis for the selection of the samples will not be sufficient unless it is the same as that to be used in the comparisons involved in the process of grading.

4. Those who use the scale with the purpose of accurate comparisons of results must be technically trained for such work, just as poultry judges, for example, find training necessary. On the other hand, such scales, though used with somewhat different results, will be valuable in ordinary schoolroom practice.

5. The scale itself should be based upon the factors to be considered in the grading.

6. Elaborate efforts to establish a zero of quality in order to have a "perfect scale" analogous to simple physical measures are quite unnecessary. Ideally, any two samples may be taken as fixed points and other samples may be defined in terms of their difference.

7. The definition of samples should be based upon the best available schoolroom practice, rather than upon an attempt to represent a progressive increase in the perfection of various factors which may or may not represent writing in its various stages of growth. The samples on the scale should represent the various degrees of excellence as they actually mature. I am inclined to believe that the factors of production need to be given more weight in the determination of scales, as some have urged. The excellence of a product is not to be determined by a consideration merely of the ratio certain factors bear to the same factors in the finished product, but by a consideration of its general scheme of growth toward that end. Moreover, it is worth while to measure the process as well as the product. In the grading of the product, then, the characteristics of the process should have consideration in the degree that desirable or undesirable features of the process are shown by peculiarities of the product.

8. The next step, therefore, is to determine and to describe in objective terms the characteristics of handwriting as found in the best practice of different systems. This would better be the work of specialists in handwriting who are willing to devote themselves faithfully to this end. On the basis of these studies the essential factors of handwriting could be selected and defined by a representative body of school men. After the objective definition of the grades of writing to be recognized, the number of actual publications known as scales may be as numerous as the different systems of writing, and yet all would have the same foundation, just as many styles of thermometers may read in Fahrenheit units. Such scales will furnish common terms for the discussion of handwriting and at the same time a fairly true picture of the thing it is really worth while to measure—namely, the progressive accomplishment of a pupil toward the acquirement of a certain ability in a particular line of activity. Meanwhile, however, let us realize that we have made progress, and let us make the best use of the tools so far available.